

Gene-Gene Interaction Analysis for the Genome-wide Association Studies

Kyunga Kim¹, Wonil Chung¹, Sung-Gon Yi¹, Junghyun Namkung², Taesung Park^{1§}

¹Department of Statistics, Seoul National University, 56-1 Shillim-Dong, Gwanak-gu, Seoul 151-742, Korea

²Interdisciplinary Program in Bioinformatics, Seoul National University, 56-1 Shillim-Dong, Gwanak-gu, Seoul 151-742, Korea

§To whom correspondence should be addressed.

Email addresses:

TP: tspark@stats.snu.ac.kr

Abstract

Gene-gene interaction is well recognized as playing an important role in understanding complex traits. Multifactor dimensionality reduction (MDR) was proposed by Ritchie, et al. (2001) to identify the multiple loci that affect disease susceptibility simultaneously. Although the MDR method has been widely used to detect gene-gene interactions in disease association studies, it cannot be easily applied to handle the whole genome association due to two limitations. First, the MDR procedure would require heavy computing in the genome-wide association research because the interaction effects are evaluated for all possible SNP combinations. Second, the evaluation is based on a cross-validation (CV) approach, and the single best SNP combination for each-order interaction is usually reported without declaring statistical significance.

We propose a simple stepwise strategy to unravel the gene-gene interactions in the framework of MDR approach for the whole genome association studies. First, we provides a way to identify a set of SNP combinations that have (statistically) significant interaction effects by developing a test statistic based on the balanced accuracy. The empirical thresholds are obtained via permutations to declare the statistical significance. Second, a stepwise procedure is developed to reduce the heavy computing due to the large search space. In the first step, the marginal effects are tested for all SNPs, and the significant effects are reported. In the second step, we test the second-order interactions only for the SNP combinations of which no or only one SNP was reported with a significant marginal effect. In the next step for testing third-order interactions, we exclude three-way SNP combinations if all SNPs were reported with significant

marginal effects and/or second-order interactions. The same strategy can be applied to the higher-order interactions. By avoiding tests for combinations of which all SNPs has been reported with significant marginal and/or any lower-order interaction effects, one can reduce a considerable amount of computing and conduct effectively an analysis of gene-gene interactions based on the MDR method in the genome-wide association studies.